

What is claimed is:

1. A method for comparing a first sequence and a second sequence, the method comprising:
 associating errors with alignments of the first sequence and the second sequence,
 comparing the alignment errors to identify the alignment having the smallest error, and,
 based on the alignment having the smallest error, computing: a first percent identity
 relative to the first sequence, and a second percent identity relative to the second
 sequence.
2. A method according to claim 1, further including determining at least one of:
 a mismatch number based on mismatches between the first sequence and the second
 sequence based on the alignment having the smallest error, and,
 an alignment number based on matches between the first sequence and the second
 sequence based on the alignment having the smallest error.
3. A method according to claim 2, where:
 the mismatches are negative matches, and,
 the matches can be at least one of perfect matches and positive matches.
4. A method according to claim 1, where computing a first percent identity relative to the first
sequence includes:
 determining an alignment number based on the matches between the first sequence and
 the second sequence based on the alignment having the smallest error, and,
 forming a ratio based on the alignment number and the length of the first sequence.
5. A method according to claim 4, where:
 the mismatches are negative matches, and,
 the matches can be at least one of perfect matches and positive matches.

6. A method according to claim 1, where computing a second percent identity relative to the second sequence includes:
 - determining an alignment number based on the matches between the first sequence and the second sequence based on the alignment having the smallest error, and,
 - forming a ratio based on the alignment number and the length of the second sequence.
7. A method according to claim 6, where:
 - the mismatches are negative matches, and,
 - the matches can be at least one of perfect matches and positive matches.
8. A method according to claim 1, further including computing a third percent identity relative to the alignment having the smallest error.
9. A method according to claim 8, where computing a third percent identity includes:
 - determining an alignment number based on the matches between the first sequence and the second sequence based on the alignment having the smallest error, and,
 - forming a ratio based on the alignment number and the length of the alignment.
10. A method according to claim 1, further including,
 - determining whether at least one of the first percent identity and the second percent identity is greater than a percent identity threshold.
11. A method according to claim 10, further including providing a percent identity threshold.
12. A method according to claim 1, further including determining at least one of:
 - a number based on the gaps in the first sequence based on the alignment having the smallest error, and,
 - a number based on the gaps in the second sequence based on the alignment having the smallest error.
13. A method according to claim 1, further including:

providing at least one database, the at least one database including at least one sequence,
and,
retrieving at least one of the first sequence and the second sequence from the at least one
database.

14. A method according to claim 1, where,
the first sequence includes at least one of: at least one polypeptide sequence and at least
one nucleotide sequence, and,
the second sequence includes at least one of: at least one polypeptide sequence and at
least one nucleotide sequence.
15. A method according to claim 1, where associating errors with alignments includes,
aligning the first sequence and the second sequence, and,
computing an error based on the number of mismatches in the alignment.
16. A method according to claim 1, where associating errors with alignments includes,
aligning the first sequence with the second sequence based on at least one insertion event
in at least one of: the first sequence and the second sequence.
17. A method according to claim 1, where associating errors includes computing a string edit
distance.
18. A method according to claim 1, where associating errors includes comparing a number of
alignment errors to an alignment error threshold.
19. A method according to claim 1, where associating errors with alignments includes,
comparing a length of the first sequence to a length of the second sequence to identify a
shorter sequence and a longer sequence, and,
aligning at least the entirety of the shorter sequence with at least a fragment of the longer
sequence.

20. A method according to claim 19, where aligning at least the entirety includes inserting at least one gap into at least one of the shorter sequence and the longer sequence.
21. A method according to claim 19, where comparing includes,
- determining that the first sequence length is equal to the second sequence length, and,
 - associating the first sequence with the shorter sequence and the second sequence with the longer sequence, and performing the aligning, and,
 - associating the first sequence with the longer sequence and the second sequence with the shorter sequence, and performing the aligning.
22. A method according to claim 19, where comparing includes,
- determining that the first sequence length is equal to the second sequence length, and,
 - associating at least one of:
 - the first sequence with the shorter sequence and the second sequence with the longer sequence, and,
 - the first sequence with the longer sequence and the second sequence with the shorter sequence.
23. A method according to claim 1, where associating errors includes aligning regardless of homology.
24. A method according to claim 1, where associating errors includes performing at least one pairwise alignment.
25. A method according to claim 1, where associating errors includes implementing a dynamic programming module for approximate string matching.
26. A method according to claim 1, further including:
- comparing the length of the first sequence with the length of the second sequence, and,

performing the alignments based on the length comparison and a percent identity threshold.

27. A method according to claim 1, further including providing at least one interface to perform at least one of: identify the first sequence, identify the second sequence, provide a percent identity threshold, and provide an alignment error threshold.

28. A method according to claim 1, further comprising outputting the first percent identity and the second percent identity.

29. A method according to claim 1, further comprising outputting the first percent identity and the second percent identity based on at least one of: a percent identity threshold and an alignment error threshold.

30. A method according to claim 1, further comprising outputting a scoring matrix associated with the first percent identity and the second percent identity.

31. A method according to claim 1, further comprising outputting data based on a comparison of the first percent identity and the second percent identity with a percent identity threshold.

32. A method according to claim 1, further comprising:

iteratively,

storing the first percent identity and the second percent identity,

retrieving at least one of a first sequence and a second sequence, and,

returning to associating errors,

to provide at least one stored first percent identity and second percent identity.

33. A method according to claim 32, where storing includes associating the first percent identity and the second percent identity with at least one of the first sequence and the second sequence.

34. A method according to claim 32, further comprising:

sorting the at least one stored first percent identity and second percent identity based on percent identity, and,
outputting the sorted first percent identity and second percent identity.

34. A method according to claim 1, further comprising:

performing in at least one parallel processing thread,
storing the first percent identity and the second percent identity, and,
retrieving at least one of a first sequence and a second sequence, and,
returning to associating errors,
to provide at least one stored first percent identity and second percent identity.

35. A method according to claim 1, where at least one of the first sequence and the second sequence includes an ASCII string.

36. A method according to claim 1, where at least one of the first sequence and the second sequence includes an identifier to a database of sequences.

37. A method according to claim 1, where the first sequence includes an identifier to a first database and the second sequence includes an identifier to a second database of sequences.

38. A method according to claim 37, where the first database and the second database are the same.